

国立国語研究所学術情報リポジトリ

表紙，目次，中扉，奥付

雑誌名	近代語コーパス設計のための文献言語研究 成果報告書
発行年	2012-10-31
シリーズ	国立国語研究所共同研究報告 ； 12-03
URL	http://doi.org/10.15084/00002760

近代語コーパス設計のための文献言語研究 成果報告書

田中牧郎・岡島昭浩・小木曾智信・小野正弘・小島聡子・島田泰子・

朱京偉・高田智和・張元哉・陳力衛・近藤明日子・須永哲矢

2012 年 10 月

近代語コーパス設計のための文献言語研究 成果報告書

目 次

本報告書の目的と概要(田中牧郎)-----	7
-------------------------	---

- (1) 本報告書の目的
- (2) 本報告書の概要
- (3) 今後に向けて

第1部 コーパスの設計

[資料選定]

1 .近代語コーパスにおける資料選定の考え方(田中牧郎)-----	13
-------------------------------------	----

- (1) はじめに
- (2) 『太陽コーパス』から近代語コーパスへ
- (3) 近代語の資料リストの作成
- (4) 資料リストの分類と資料選定の考え方 明治前期・中期を例に
- (5) 資料選定の実施に向けて
- (6) おわりに

2 . 電子化が望まれる近代語資料探索 日本語史を研究する大学院生の報告から (岡島昭浩・森勇太・金囀泳・竹村明日香・坂井美日)-----	27
--	----

- (1) 趣旨
- (2) 提案されたもの
- (3) 例
- (4) まとめ

[文字処理]

3 .近代語文献を電子化するための文字セット(高田智和)-----	36
-------------------------------------	----

- (1) はじめに
- (2) 『太陽コーパス』の文字処理
- (3) 『太陽コーパス』の JIS X0213 による再符号化
- (4) おわりに

4 .近代語文献を電子化するための異体字処理(須永哲矢)----- 65

- (1) はじめに
- (2) JIS X0213 文字集合と包摂規準
- (3) 『明六雑誌』漢字処理上の問題
- (4) 近代語コーパスのための文字処理方針
- (5) 『明六雑誌』漢字字形処理方針
- (6) 追加包摂規準・別字代用一覧
- (7) JIS X0213 文字集合 / 追加包摂 / 別字代用の検証
- (8) 最終的に「≡」表示となる外字一覧
- (9) 今後の展望

[形態素解析]

5 .近代語テキストの形態素解析(小木曾智信)-----83

- (1) はじめに
- (2) 近代語の形態素解析
- (3) 近代文語 UniDic
- (4) 近代語コーパスへの形態論情報付与(『明六雑誌』の場合)
- (5) おわりに

6 .近代語コーパスのための形態論情報付与規程の整備(須永哲矢・近藤明日子)----- 93

- (1) 近代語コーパスでの言語単位
- (2) 近代語での単位認定の問題点と、その処理方針
- (3) 今後の課題
 - 資料 1 : 仮名表記される外来語の語形の定め方
 - 資料 2 : 出現形「に」の品詞判別基準

[モデルコーパス]

7 .『明六雑誌コーパス』の仕様(近藤明日子・田中牧郎)-----118

- (1) はじめに
- (2) 『明六雑誌』を選ぶ理由
- (3) 文字入力の基本仕様
- (4) XML タグセット
- (5) コーパスの公開形式

8 .『明六雑誌コーパス』の語彙量(近藤明日子)----- 144

- (1) 本稿の目的

- (2) 凡例
- (3) 語彙量の報告

第2部 コーパスの活用

[語彙研究]

9 . 明治後期から大正期の語彙のレベルと語種

- 『太陽コーパス』の形態素解析データによる (田中牧郎)----- 153
- (1) はじめに
- (2) 『太陽コーパス』への「近代文語 UniDic」の適用
- (3) 『太陽コーパス』の語種比率
- (4) 『太陽コーパス』の語彙のレベル分け
- (5) レベルの変動による類型化
- (6) レベルから見た和語の特徴
- (7) レベルから見た漢語の特徴
- (8) おわりに

10 . 文献資料内漢語の階層化 『明六雑誌』の漢語をめぐって (小野正弘)-----169

- (1) はじめに
- (2) 具体的手順
- (3) 分析結果
- (4) おわりに

11 . 『明六雑誌』の一人称代名詞(近藤明日子)----- 181

- (1) はじめに
- (2) 『明六雑誌コーパス』の概要
- (3) 分析対象とする語の抽出とその度数の概観
- (4) 語と後続助詞との対応関係
- (5) 連体用法における語と被修飾体言との対応関係
- (6) 主な語の特徴
- (7) おわりに

[文法研究]

12 . 近代書き言葉における文語助動詞から口語助動詞への推移

- 『太陽コーパス』の形態素解析データによる (田中牧郎)----- 191

(1) はじめに	
(2) 『太陽コーパス』における文語体と口語体	
(3) 各年次 5 万レコードの調査	
(4) 助動詞の頻度	
(5) 断定の助動詞の分析	
(6) おわりに	
13 . 近代語に探る 終止形準体法 の萌芽的要素 (島田泰子) -----	201
(1) はじめに	
(2) 終止形準体法 について	
(3) コーパスを利用した用例採集	
(4) 実例から (気付かれる点)	
(5) おわりに	
14 . 近代の地方出身作家の助詞の用法について	
宮澤賢治と濱田廣介 (小島聡子) -----	211
(1) はじめに	
(2) 宮澤賢治と濱田廣介	
(3) コーパスの利用について	
(4) 格助詞の用法	
(5) 接続助詞 (接続詞)	
(6) 副助詞等について	
(7) 今後の課題	
[日中韓対照研究]	
15 . 『太陽コーパス』における漢文系複合辞の使われ方 (朱京偉) -----	221
(1) はじめに	
(2) に基づく / 基於 (基于)	
(3) に関する / 關於 (关于)	
(4) に対する / 對於 (对于)	
(5) に由る / 由於 (由于)	
(6) と認め / 認為 (认为)	
(7) と成る / 成為 (成为)	
(8) と視る / 視為 (视为)	
(9) まとめ	

16 . 日中の比較語史研究（陳力衛）-----	237
(1) 問題提起	
(2) 中国語資料を手掛かりに	
(3) 『日本国語大辞典』の初出例	
(4) 近代資料とは何か	
(5) 日中言語交流の時間的幅の設定	
(6) 終わりに	
17 . 近代対訳コーパスにおける日韓語彙の諸相	
文体の異なる対訳コーパスの比較を通して（張元哉）-----	247
(1) はじめに	
(2) 調査資料と調査方法	
(3) 日韓の語彙量の対照	
(4) 日韓の語種構成の対照	
(5) 日韓の品詞構成の対照	
(6) 日韓の語構成の対照	
(7) おわりに	
共同研究発表会開催記録-----	259

第 1 部 コーパスの設計

第 2 部 コーパスの活用

執筆者一覧

田中 牧郎	(国立国語研究所言語資源研究系准教授)
岡島 昭浩	(大阪大学大学院文学研究科教授)
小木曾智信	(国立国語研究所言語資源研究系准教授)
小野 正弘	(明治大学文学部教授)
小島 聡子	(岩手大学人文社会科学部准教授)
島田 泰子	(二松学舎大学文学部教授)
朱 京偉	(中国・北京外国語大学教授)
高田 智和	(国立国語研究所理論・構造研究系准教授)
張 元哉	(韓国・啓明大学副教授)
陳 力衛	(成城大学経済学部教授)
近藤明日子	(国立国語研究所コーパス開発センタープロジェクト奨励研究員)
須永 哲矢	(国立国語研究所コーパス開発センタープロジェクト奨励研究員)
金 曙泳	(韓国・高麗大学校言語情報研究所)
坂井 美日	(大阪大学大学院博士後期課程学生)
竹村明日香	(大阪大学大学院博士後期課程学生)
森 勇太	(日本学術振興会特別研究員)

国立国語研究所共同研究報告 12-03

近代語コーパス設計のための文献言語研究 成果報告書

2012 年 10 月 31 日発行

著者 田中牧郎・岡島昭浩・小木曾智信・小野正弘・小島聡子・島田泰子・
朱京偉・高田智和・張元哉・陳力衛・近藤明日子・須永哲矢

発行 大学共同利用機関法人人間文化研究機構国立国語研究所

〒190-8561 東京都立川市緑町 10-2

電話 042(540)4300 (代表)

<http://www.ninjal.ac.jp/>

©国立国語研究所

ISBN 978-4-906055-23-4

ISSN 2185-0127

Study on Documents and Meta-languages
for Designing a Corpus of Modern Japanese

Makiro Tanaka, Akihiro Okajima, Toshinobu Ogiso,
Masahiro Ono, Satoko Kojima, Yasuko Shimada,
Jingwei Zhu, Tomokazu Takada, Wonjae Chang,
Liwei Chen, Asuko Kondo, Tetsuya Sunaga

October 2012